TATJANA CHERNENKO

Applied AI & Data Scientist | Speech AI, NMT, Agentic AI, KG | Applied Research Leadership & Advisory
tatjana.chernenko@gmail.com | LinkedIn: https://www.linkedin.com/in/tatjana-chernenko/ |
GitHub: https://github.com/TatjanaChernenko |

PROFESSIONAL SUMMARY

- Applied AI Scientist with 20 years of professional experience, including ~10 years in AI, NLP, and multilingual language technologies, and 7+ years of sole or primary **end-to-end ownership of ambiguous enterprise AI initiatives** at SAP.
- Combines hands-on applied **research**, evaluation **architecture**, and AI system design with principal-level **leadership** across **research-to-production delivery**.
- Built and owned the AI research-to-production dimension of enterprise-scale language and speech platforms, including SAP's multilingual retrieval systems and SAP's Speech AI capability launched from zero to production; co-designed enterprise Agentic AI workflows, Knowledge Graph and RAG architectures, data products, and architecture of GenAI/ML-based systems.
- Particular strengths include translating state-of-the-art AI advances into enterprise-grade systems and architectural ecosystems, originating and steering complex multidisciplinary research, and providing AI leadership across research-to-production environments.

SELECTED IMPACT HIGHLIGHTS

- **Built SAP's Speech AI capability from zero to enterprise production** as the sole AI expert in a startup-style team, defining the applied research direction, evaluation architecture, quality-improvement methodology, creating enterprise speech data pipelines and SAP-specific adaptation path across multilingual speech workflows under entwerprise legal constraints; delivered up to **20% WER quility improvement** and **~10% MIM uplift** over external solutions.
- **Owned SAP applied research agendas and innovation** across Speech AI, multilingual NLP, and agentic AI, leading international academic collaborations across **Brazil, Japan, India, and Germany**, co-designing novel AI architectures, mentoring researchers, and building a track record of **granted US patents and publication output, including accepted LREC-COLING 2026 paper.**
- **Served as primary AI owner for SAP's multilingual translation retrieval platform**, driving the shift from keyword-based to semantic retrieval across **40+ languages**, **2,000+ language combinations**, and millions of verified translations, while establishing the evaluation and quality-measurement backbone that improved proposal quality and supported enterprise-scale customer rollouts.
- **Provided principal-level AI direction across platforms, products, and emerging AI initiatives**, advising leadership, product, and engineering stakeholders on feasibility, architecture, evaluation, and risk across Speech AI, agentic AI, RAG, knowledge graphs, GenAI/ML-based systems and data products, including contribution to **100+ internal AI automation ideas** and cross-functional execution under governance, licensing, and supplier constraints.
- Earlier career includes **entrepreneurial leadership as CEO, analytical and management experience** in international business environments.

CORE EXPERTISE

Speech AI (ASR, TTS, speech translation) • NMT • Multilingual NLP • AI Evaluation & Benchmarking Frameworks • Research-to-Production • Enterprise AI Architecture • Retrieval & Semantic Search • Low-Resource / Terminology-Aware Adaptation • GenAI & ML architectures • RAG • Agentic AI • Knowledge Graphs • Data Products & Pipeline Design • AI Quality Improvement • Quality Governance • Complex Evaluation & Benchmarking Frameworks from scratch (Speech AI, NMT, Agentic AI, Retrieval) • Model Optimization • LLM judge auditing • Vector Retrieval • Fine-tuning • Data Processing • Cloud Infrastructure • Applied Research Leadership • AI Innovation

PROFESSIONAL EXPERIENCE

Lead AI Scientist / Speech AI | SAP SE 2019 – Present

Sole AI expert across several enterprise-scale platforms; owned AI direction, evaluation architecture, applied research, and production enablement.

Enterprise Speech AI — Evaluation, Research & Production (2024–2026)

- **Launched SAP's Speech AI (ASR, TTS and Speech Translation) from zero** - no prior product or benchmarks existed. Defined and implemented entire research-to-production AI direction as sole AI specialist on a 5-person team (3 SWE + 1 PO). **Product shipped to enterprise production, April 2026.**
- Built evaluation framework end-to-end: ~90 datasets, ~12 languages, 100+ dimensions. Implemented STT, TTS, E2E evaluation pipelines (coming soon as scientific papers). Integrated automated metrics with human evaluation; built decision logic linking offline metrics, robustness checks, and improvement prioritisation into a reusable quality-governance layer.
- Quality improvements for 8 languages: ASR: OVKWS-architecture for terminology injection, noise robustness, LLM-based post-editing, fine-tuning. TTS: text engineering, Grapheme-to-Phoneme. E2E: NLP- and LLM-based pre- and post-processing. **Delivered up to 20% WER improvement on the SAP domain** and ~10% MIM improvement on selected terminology over external solutions. Voice cloning.
- Designed patented ASR stabilisation concepts. Co-authored patents*: "Adaptive Fidelity Pipeline for Minimising Hallucinations and Skipped Content in STT Systems"* (pending, 250089US01).
- Enabled enterprise AI execution in compliance-sensitive environments by translating legal, licensing, procurement, data-protection, and supplier constraints, de-risked initiatives involving highly sensitive enterprise audio data.

Applied Research: Specialised Vocabulary & Low-Data ASR (UNISINOS, 2025–2026)

- Originated and led research on customer-specific terminology injection for enterprise ASR. Co-designed novel **open-vocabulary keyword spotting (OVKWS) architecture** using TTS-generated synthetic audio, cross-attention for acoustic matching and hard-negative sampling.
- Led international collaboration (AI Scientist + 2 Linguists + ~3 Professors); evaluated and integrated the method into the productive solution.
- **Achieved +25% MRR (EN) and +133% MRR (PT)** on company-domain benchmarks. Accepted **LREC-COLING 2026 paper**: *"A Dataset for Evaluating ASR on Specialised Vocabulary"* and an internal *SAP paper covering the architecture and training approach (confidential due to SAP constraints).*

Speech AI for Robotics (2025–2026)

- **Frontier AI Research (Speech AI for Robotics):** Contributed to international applied research on Speech AI for robotics and voice-enabled intelligent systems (collaboration involving SAP Japan, SAP Research & Innovation in Potsdam, TRUSCO Nakayama, AIMBO Robotics, and Nagoya University's Kawaguchi Laboratory, adding frontier-facing research breadth across robotics, embodied AI, warehouse automation, and S/4HANA-connected intelligent systems).

Japanese Enterprise TTS Optimisation (2024–2025)

- Owned problem framing for terminology-focused TTS improvement in Japanese in settings where extensive fine-tuning was impractical due to data sensitivity, compliance overhead, and cost. Mentored a student in Japan.
  Up to 15% WER improvement. Co-authored patent *"Enhancing Pronunciation Accuracy and Naturalness in Japanese TTS without Extensive Fine-tuning"* (currently under review at SAP).

SAP-Wide Speech Data Product Architecture (2024–2026)

- Architected SAP-wide multilingual speech data product: data sourcing and creation, 20+ stage processing pipeline, 100+ metadata dimensions; infrastructure: AWS S3 + SAP DataSphere.

Agentic AI, GenAI/ML-based Systems, RAG & Knowledge Graphs (2024–2026)

- Led AI-side feasibility, architecture and research for Query Reduction (GlobalizeMe) and SAP Agentic AI POCs. Co-screened 100+ AI automation research-to-production capabilities in collaboration with leadership, architects and stakeholders. AI expert at leadership offsite (Brazil) and AI workshops.

Terminology Intelligence & Patent Innovation (2019–2025)

- Analysis of 12M+ multilingual term pairs. 3 granted US patents (*semantic domain assignment, disambiguation, abbreviation resolution*).

Enterprise Retrieval Platform: SAP Translation Hub MLTR 2.0 (2019–2024)

- Primary AI owner: 40+ languages, millions of translations, 2,000+ language combinations. Drove keyword-to-semantic retrieval transition. **Up to 10% improvement in proposal quality.** Evaluation architecture: BLEU, COMET, ChrF, embedding similarity; 36+ reports per cycle. Enterprise rollouts: 80k+ text elements, 25k+ objects. Adopted by Siemens Healthineers, Daimler Truck (93% reduction in manual effort, 99%+ accuracy, 1,250 hours saved).
- Led cross-team standardisation of data interchange across five core teams, defining common formats and a shared processing library that improved interoperability and reduced communication overhead.

Applied AI Developer / Data Scientist | SAP SE, Germany 2018 – 2019

- Built a task-oriented conversational AI system for technical support from scratch: dialogue management, deep learning, and semantic search. Full end-to-end ownership.
- End-to-end data pipeline for complex enterprise data (unstructured support tickets, HTML, XML), incl. advanced pre-processing, AI-based intent classification, filtering, and feature extraction.
- Architecture: Semantic retrieval on vectorized KB (domain fine-tuned Word2Vec embeddings enriched with one-hot encoded metadata)

Data Scientist / ML Engineer | Empolis Information Management, Germany 2018

- Information extraction, text mining, NLP-driven scientific paper search platform for the pharmaceutical domain.

Computational Linguist | Spiegel Institut, Germany 2016

- Spoken-language processing for autonomous driving: NLP for vehicle communication (STT, TTS).

CEO / Founding Partner | Quintessentially Estates, Ukraine / United Kingdom 2006 – 2015

- International real-estate leadership. Strategy, stakeholder management, analytics, cross-functional decision-making, customer-facing role.

EDUCATION

Doctoral Research Portfolio in Preparation - PhD by Publication Route [2026]

Preparing a PhD by Publication prospectus on robust enterprise Speech AI.

BSc, Computational Linguistics (AI focus) | Heidelberg University [2014-2018]

PhD track offered by supervising professor (Pr. Dr. Riezler); accelerated Master's track also offered; declined due to SAP recruitment.

Institute of Hospitality Management in Prague [2005-2009]

LANGUAGES

English (fluent), German (fluent), Russian (native), Czech (listening comprehension)

SELECTED PATENTS & PUBLICATIONS

Granted US Patents

- **US-12518105-B2 (Jan 2026):** "Semantic Domain Assignment Referencing Governance Domains and Term Databases"
- **US-12386820-B2 (Aug 2025):** "System and Method Performing Terminology Disambiguation"
- **US-12067370-B2 (Aug 2024):** "Detection of Abbreviation and Mapping to Full Original Term"

Pending / Under Review

- **250089US01 (Aug 2025):** "Adaptive Fidelity Pipeline for Minimizing Hallucinations and Skipped Content in STT Systems"
- **Under review, 2026:** "Enhancing Pronunciation Accuracy and Naturalness in Japanese TTS without Extensive Fine-tuning"

Publications & Academic Research

- **LREC-COLING 2026 (May 2026):** "A Dataset for Evaluating ASR on Specialised Vocabulary" — evaluation methodology and benchmark design for specialised-vocabulary robustness.
- SAP papers covering AI architecture and training approaches (confidential due to SAP constraints).
- **NLG from Structured Inputs (Heidelberg, 2018):** Encoder–decoder pipeline (feed-forward + LSTM) for image description generation. MS COCO, V-COCO, COCO-a.
- **CHERTOY: Word Sense Induction (Heidelberg, 2017):** Unsupervised semantic clustering. sense2vec + MeanShift, 40 ablation experiments. RI 0.37→0.68, Jaccard 0.14→0.62.

INTERNATIONAL RESEARCH COLLABORATIONS

15+ initiatives: UNISINOS (Brazil, specialised vocabulary ASR) · **IIIT Bangalore** (Indic ASR) · **Nagoya University / TRUSCO Nakayama / AIMBO Robotics** (Japan, Speech AI for robotics)